An Application of the Runs Test to Test for Randomness of Observations Obtained from a Clinical Survey in an Ordered Population

Mohamad Adam Bujang, **Fatin Ellisya** Sapri

Clinical Research Centre, Sarawak General Hospital, Ministry of Health, Malaysia

Abstract

Runs test is a statistical procedure which determines whether a sequence of data within a given distribution have been derived with a random process or not. It may be applied to test the randomness of data in a survey that collect data from an ordered population. This article illustrates on method to perform a runs test and explains the rationale for performing it by providing some examples of how this test can be applied. The aim of this article was to describe on ways to use the runs test in a clinical survey from an ordered population to determine the degree of randomness in the sequence of subjects who are recruited within a sample obtained from the whole population. Clinical survey that involves an ordered population usually collects data from subjects who have been recruited by a consecutive sampling method. Therefore, this study recommends that the degree of randomness in the sequence of selected variable(s) obtained from consecutive sampling is necessary to be tested from a pilot study to ensure random data collection in the study.

Keywords: random, runs, sample, sampling, survey

Introduction

The terms 'random' and 'non-random' are commonly used to describe whether data have been obtained using a random or a non-random sampling procedure (1). For a clinical survey, a random sample of data will be collected if the probability sampling method has been adopted (2). On the other hand, although it is still possible for non-probability sampling to generate a random data, there is, however, no assurance that the sequence of data generated will be random because there is no mechanism to ensure random selection.

The present article emphasised sample selection from clinical survey in an ordered population. An ordered population is based on a 'first come, first served' manner. This type of population is usually occurring from a queue system such as when the patients are queuing while waiting to be treated by physicians or the cars are queuing for making an order at fast food restaurants. Among those sampling techniques under non-probability sampling, consecutive sampling is the commonest and easiest technique for an ordered population. Since consecutive sampling is one of the non-probability sampling techniques, it is necessary to check whether or not a random data has been collected from the process of recruitment. This is because statistical inferences drawn from a population using inferential statistics can only be valid if sample data have been collected without bias (3).

The Concept and Application of Runs Test

A run is a sequence of events of a certain type preceded and followed by occurrences of the alternate type or by no events at all. A sample with too many or too few runs suggests that the

sample may not be random. The runs test is a statistical test to determine whether random selection has been made in the process of sample selection from an ordered population. The runs test is a type of non-parametric test and hence, there is no need for the assumption of a normal distribution to hold true. It has been stated that for one sample test, the variable must be dichotomous while for two-sample test, the two samples should be mutually independent (4). This article aimed to describe the procedures in performing the runs test, and then illustrate how this test can be applied by using some real-life examples. The application of runs test in testing the randomness in the selection can become an evidence to show that there is no bias in the selection process.

Hypotheses

There are several hypotheses can be tested using the runs test. The researcher has to select the correct hypothesis statement depending on which hypothesis the researcher wishes to test. The various hypothesis statements are as follows:

1. Two-sided

    H_0: The occurrence of pattern for the two types of the observations is determined by a random process.

    H_1: The occurrence of pattern for the two types of the observation is not determined by a random process.

2. One-sided

    H_0: The occurrence of pattern for the two types of the observations is determined by a random process.

    H_1: The occurrence of pattern for the two types of the observation is not determined by a random process because there are too few runs.

3. One-sided

    H_0: The occurrence of pattern for the two types of the observations is determined by a random process.

    H_1: The occurrence of pattern for the two types of the observation is not determined by a random process because there are too many runs.

Say a researcher observes a sequence of 20 patients attending at clinic A. The researcher would now like to determine whether the sequence is following a random sequence based on gender (M= Male and F= female).

F F M M F M F F F F M F F M M M M M F F

First of all, the researcher has set the hypothesis with two-sided test. The hypothesis with two-sided test can now be re-defined to make it more precisely reflecting the specific scenario for the hypothesis testing, as shown below:

H_0: The pattern of sequence in terms of gender attended at clinic A is arranged in random manner.

H_1: The pattern of sequence in terms of gender attended at clinic A is arranged in non-random manner.

Next is to count the number of observations for female, male and the number of runs for this series:

A run is defined as a series of consecutive observations' values. In this example, the researcher shall count number of consecutive for both M and F. The calculation for the runs is as follow;
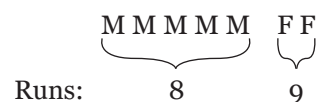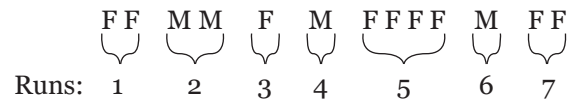
Therefore, number of female (n_1) = 11, number of male (n_2) = 9 and runs (r) = 9.

Test Statistics

Next, is to calculate the test statistic. Test statistic is a standardised value that is calculated from sample data for a hypothesis test which measures the agreement in between the sample data. The calculated value of the test statistic is used to compare the data obtained from the experimental conditions with the data expected to be obtained if the null hypothesis is valid.

Therefore, the test statistic is used to determine whether or not the null hypothesis should be rejected. The methods of calculation for both test statistics and critical values are different between a small sample and a large sample.

Generally, test statistics for small sample will be based on the number of runs, while approximation technique will be applied for large sample. Besides, critical values for small sample will be obtained from a runs test for randomness table while a formula will be used to generate critical values for large sample. The sample sizes are considered as small when both n_1 and n_2 are 20 or less (1, 5–6). In other words, it is recommended that, for sample size more than 20 to apply approximation method.

Therefore, the following example will be considered to be a small sample because the observations are 20. Thus, the test statistics for a small sample is considered equivalent to the number of runs. Referring to previous example, there are 9 runs in event 1 and event 2, therefore test statistics is $r = 9$.

Critical Value for Small Samples

The critical value can be obtained from table by Swed and Eisenhart (1). Hence, from the table, with $n_1 = 11$ and $n_2 = 9$, the upper critical value (Uc) obtained was six and the lower critical value (Lc) obtained was 16.

Decision Rule

The following is the decision rule for a small sample based on the different types of hypothesis statements:

1. Two-sided

   Reject H_0:

   $r \leq$ Lc or $r \geq$ Uc

2. One-sided (H_1 has too few runs)

   Reject H_0:

   $r \leq$ Lc

3. One-sided (H_1 has too many runs)

   Reject H_0:

   $r \geq$ Uc

From this example, since $6 \leq r = 9 \leq 16$, hence, the decision is to accept H_0. Therefore, we can conclude that there is enough evidence to claim that the sequence in the gender of patients attending clinic A had been selected in a random manner.

Approximation Technique for Large Samples

The next discussion is to describe on how to perform the runs test for large samples of numerical (or continuous) data. Let's take an example of HbA1c level obtained from patients with diabetes mellitus in clinic B. The data-set are as follow:

7.7, 10.0, 8.8, 12.1, 7.4, 6.1, 6.1, 6.1, 8.0, 8.8, 6.1, 10.9, 5.0, 12.6, 11.7, 11.5, 6.4, 8.9, 7.2, 7.2, 6.5, 10.3, 7.9, 7.8, 14.0, 6.8, 7.7, 7.8, 6.7, 7.7

For numerical data, researcher will need to first categorise it into two groups based on certain cut-off points. The cut-off points can be obtained from mean, median, mode or by custom (as determined by the researcher). Cases with values less than the cut-off point are categorised in one group while cases with values more than the cut-off point are categorised into another group. In this example, the researcher will use the custom cut-off point of 7, HbA1c < 7 as group 1 indicating good control and HbA1c ≥ 7 as group 2 which indicates poor control. The data were grouped as follows:

2 2 2 2 2 1 1 1 2 2 1 2 1 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 1 2

Next, the researcher need to set hypotheses for this analysis. Let's say the null hypothesis is the pattern of sequence in terms of HbA1c level (poor or good control) observed at clinic B is arranged in random manner. Then, the next step is to calculate the test statistic. For a large sample, the test statistic can be calculated by using an approximation of the normal distribution via the following formula (7):

$$z = \frac{r - \mu_r}{\sigma^r} \qquad (1)$$

where:
r is the number of runs;
μ_r is the expected number of runs; and
σ_r is the standard deviation of the number of runs.

The values of μ_r and σ_r are computed as follows:

$$\mu_r = \frac{2n_1 n_2}{n_1 + n_2} + 1 \qquad (2)$$

$$\sigma_r = \sqrt{\frac{(2n_1 n_2)(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}} \qquad (3)$$

Let's calculate for mean and standard deviation of the runs:

$$\mu_r = \frac{2(9)(21)}{9 + 21} + 1$$
$$= 13.6$$

$$\sigma_r = \frac{[2(9)(21)][2(9)(21) - 9 - 21]}{(9 + 21)^2 9 + 21 - (1)}$$
$$= \sqrt{5.04}$$
$$= 2.245$$

Test statistics:

$$z = \frac{15 - 13.6}{2.245}$$
$$= 0.624$$

After that, the researcher need to find the critical value for large sample. So, the calculation for the critical value is as follow:

$$\text{Critical value (two tail)} = z_1 - \frac{\alpha}{2}$$
$$= z_1 - \frac{0.05}{2}$$
$$= z_{0.975}$$
$$= 1.96$$

At the 5% significance level, a test statistic with an absolute value greater than 1.96 indicates non-randomness (reject H_o). For this example, since the test statistics < 1.96, hence the analysis accept H_o. Therefore, we can conclude that there is enough evidence to claim that the sequence of subjects taken for HbA1c was in a random manner.

It is recommended to use z-score while handling a large sample or when the number of observations is more than 20. This is because the distribution of the observed number of runs would approximately follow the normal distribution that has a mean of zero and variance of one (5). If researchers want to do one-tailed runs test (for the purpose of detecting many runs), then researchers need to compare the

z-score with upper tail critical value. Although the steps to obtain the critical values for lower tail and upper tail (for a one-tailed runs test) are the same as those for a two-tailed runs test, however different formula should be applied for one-tailed runs test, as shown below:

$$\text{Critical value (upper tail)} = z_1 - \alpha$$
$$= z_1 - 0.05$$
$$= z_{0.95}$$
$$= 1.645$$

For lower-tail critical value, just simply add negative sign on the critical value.

Using the Runs Test to Test for Randomness of Observations Obtained from a Clinical Survey of an Ordered Population

There is a risk of selection bias when collecting a sample using consecutive sampling for an ordered population. Hence, it is strongly recommended to provide evidence with a runs test to demonstrate that the sequence of data is random. The runs test is useful for sampling situations in which there is an ordered sampling frame, such as in the queue system or where there is an interval between two consecutive observations (such as the presence of a specified period of waiting time between them). This is where consecutive sampling method which is based on the principle of 'first-come-first-served' commonly adopted.

In consecutive sampling, the sample collection is performed by screening each subject consecutively and including every subject who meets all inclusion criteria until the required sample size has been achieved (8). Unlike convenience sampling where the sample selection is selected based on the sample that is accessible by the researchers irrespective of whether it is a consecutive selection or not (2). Both are non-probability sampling methods and hence, there will always be a possibility for selection bias to occur (2).

Example, a researcher is conducting a research study to measure quality of life among patients with end stage renal disease (ESRD) who are attending routine follow-up in a clinic. In this scenario, the population is arranged in order and therefore, one of the easiest non-probability sampling techniques in recruiting the patients is by using the consecutive sampling

method. Assume that the quality of life was hypothesised to be decreasing as the respondents get older (9), so if the researcher recruits the sample by using consecutive sampling and the sample yields majority of younger respondents in the sample, so the measurement for quality of life may not be representative to the entire patient population. Therefore, this scenario explains on the situations where it is necessary to test the randomness in the selection process.

However, it is not necessary to test whether there is random sequence for all variables because only those relevant variables must be randomly sampled. For the measurement of quality of life among ESRD patients, a good starting point shall be to evaluate whether (or not) the age of the sample of patients who have been recruited for a clinical survey are in a random sequence. In addition, it will not be necessary to determine that all the samples collected are random. This is because the most important requirement is to ensure that the sample collection process is random.

Hence, this article proposes that the runs test has to be performed during a pilot study to ensure whether there is a random order in the age of those patients coming to the clinic. It is not recommended to perform the runs test in a full-scale research because the *P*-value will then become too small if the sample size is large, which will make it unnecessary to conduct the test (since the small *P*-value is resulting from a large sample size, and is not due to the presence of a real statistical significance). So, this study recommends to collect only a small sample (20 to 30 subjects) to determine whether there is a random sequence of the sample collected for a clinical survey or not.

Before performing the runs test, it is first necessary to get a clear understanding of the particular aspect of the target population. For example, says the researcher has already known that the prevalence of older people with ESRD is much higher than that of younger people, and so the order for arrangement of patients' age may not be random as it is more likely for the first 10 or 20 patients collected within one sample to be all elderly. In this case, it is expected that the order of patients' age to be likely not random. In this situation, it is necessary to collect a larger sample for pilot study such as 30 to 50 subjects per sample.

To conduct a pilot study using a large sample containing between 30 and 50 subjects can be difficult for certain surveys, hence, there

is an alternative by using patients' age from the past data. This is feasible if the clinic has kept a database which reports the history of patients' notification by the clinic in a 'first-come-first-served' manner. By retrieving past data from a clinic, it is feasible for us to determine the random order of patients' data, provided that there is no difference in the practice of recording patients' data between the past and the present. By doing so, researchers can save both time and cost of the study.

If it has been subsequently found that the selection process using the consecutive sampling method is not random, then further investigation must be conducted to determine whether or not the selection bias can be due to the presence of a consecutive pattern which exists in the sequence of subjects recruited for a sample. This consecutive pattern may exist because of prior arrangements made at the clinics such as the treatment of severe cases (who are also usually older patients) during the morning and the treatment of less severe cases (who are also usually younger patients) in the afternoon. If this is occurring, it is now necessary to take corrective action to overcome the lack of random sequence of sample data collection by revising the study design including the selection of sampling technique.

Another proposed solution to address the above problem is to recruit a larger sample for the clinical survey. Previous studies have found that a survey which aims to recruit a minimum sample size of 300 subjects will be very likely to provide the estimates derived from the samples which closely mimic those obtained from the real population parameters (10–11). In other words, it nullifies the presence of any consecutive patterns within the sample. Moreover, a large sample size obtained will make it more likely for the estimates derived from the sample to be closer approximation to the real population parameters, which is in line with the concept of central limit theorem (12–13).

As for conclusion, this study recommends that the randomness of observations or data derived from consecutive sampling need to be assessed from a pilot study by using runs test before collecting data in the fieldwork phase of the study. This is to ensure no bias is implemented in the process of subject recruitment. Besides, runs test can also be applied to determine whether or not there is a consecutive pattern in the sample recruitment before implementing systematic sampling.

Corrective actions need to be made if non-random pattern is observed such as re-designing the sampling technique or revising the inclusion and exclusion criteria of sample recruitment.

Acknowledgements

Authors' Contributions

Conception and design: MAB
Analysis and interpretation of the data: MAB, FES
Drafting of the article: MAB, FES
Critical revision of the article for important intellectual content: MAB, FES
Final approval of the article: MAB, FES

Correspondence

Dr Mohamad Adam Bujang
BSc (Hons) Statistics (Universiti Teknologi MARA),
MBA (Universiti Teknologi MARA),
PhD (Universiti Teknologi MARA)
Clinical Research Centre, Sarawak General Hospital,
Ministry of Health, Jalan Tun Ahmad Zaidi Adruce,
Kuching, Sarawak, Malaysia.
Tel: +6082 276820
Fax: +6082 276823
E-mail: adam@crc.gov.my

References

1. Swed FS, Eisenhart C. Tables for testing randomness of grouping in a sequence of alternatives. *Ann Math Statist*. 1943;**14(1)**:66–87. https://doi.org/10.1214/aoms/117731494

2. Levy PS, Lemeshow S. *Sampling of populations: methods and applications*. 4th ed. New York: John Wiley & Sons, Inc; 2008. https://doi.org/10.1002/9780470374597

3. Deaux RD, Velleman PF. *Introduction to statistics*. 3rd ed. Addison Wesley; 2008.

4. Corder GW, Foreman DI. *Nonparametric statistics; step-by-step approach*. 2nd ed. New Jersey: Wiley; 2009. https://doi.org/10.1002/978H18165881

5. Spiegel MR. *Theory and problems of probability and statistics*. New York: McGraw-Hill; 1992.

6. NIST/SEMATECH. *E-handbook of statistical methods*. U.S. Department of Commerce; 2013. Available from: http://www.itl.nist.gov/div898/handbook/

7. Bradley JV. *Distribution-free statistical tests*. Englewood Cliffs, New Jersey: Prentice-Hall; 1968.

8. Bowers D, House A, Owens D. *Getting started in health research*. New Jersey: Wiley, 2011. https://doi.org/10.1002/9781444341300

9. Liu WJ, Musa R, Chew TF, Lim CTS, Morad Z, Bujang A. Quality of life in dialysis: a Malaysian perspective. *Hemodial Int*. 2014;**18(2)**:495–506. https://doi.org/10.111/hdi.12108

10. Bujang MA, Ghani PA, Zolkepali NA, Selvarajah S, Haniff J. A comparison between convenience sampling versus systematic sampling in getting the true parameter in a population: explore from a clinical database. The Audit Diabetes Control Management (ADCM) registry in 2009. ICSSBE 2012–Proceedings of the International Conference on Statistics in Science, Business and Engineering: Empowering Decision Making with Statistical Sciences; 2012 Sep 10–Sep 12; Langkawi, Kedah, Malaysia; 2012. pp 499–503. https://doi.org/10.1109/ICSSBE.2012.6396615

11. Bujang MA, Sa'at N, Joys AR, Ali MM. An audit of the statistics and the comparison with the parameter in the population. *AIP Conf Proc*. 2015;**1682(1)**:050019. https://doi.org/10.1063/1.4932510

12. Brosamler GA. An almost everywhere central limit theorem. *Math Proc Cambridge Philos Soc*. 1988;**104(3)**:561–574. https://doi.org/10.1017/S0305004100065750

13. Lacey MT, Philipp W. A note on the almost sure central limit theorem. *Stat Probabil Lett*. 1990;**9(3)**:201–205. https://doi.org/10.1016/0167-7152(90)90056-D